

Service Business (2007) 1:119–139
DOI 10.1007/s11628-006-0009-4

ORIGINAL PAPER

Service value chains and effects of scale

Bart Nooteboom

Received: 12 July 2006 / Accepted: 15 September 2006 / Published online: 21 December 2006
© Springer-Verlag 2006

Abstract This paper generalizes Porter's notion of the value chain for the analysis of service industries. The generalization entails that the flow and the physical transformation and assembly of goods that are characteristic of manufacturing are generalized into flows and transformation of data and flows and transformation of the physical and mental condition of people that are characteristic of many service industries. Utility is generalized from the utilities of forms and function of goods, characteristic of manufacturing, to the utilities of time, place, convenience, speed, safety, entertainment, physical and mental well-being, knowledge and mental capacity, funding and assurance. The analysis yields a categorization of industries according to central features of the value adding process. Here, the analysis is used to identify sources of (in)efficiency of scale, scope and experience, along the value chain.

Keywords Value chain · Service strategy · Effects of scale

1 Introduction

One feature of industrial dynamics, i.e. structural change in different industries, is that economies of scale lead to concentration of sales in large units of production or firms. Thus, for an understanding of industrial dynamics, it is important to understand economies of scale. Increase of scale and concentration affect productive, allocative and dynamic (innovative) efficiency. Here

B. Nooteboom (✉)
Faculty of Economics and Business Administration, Organization and Strategy,
Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
e-mail: b.nooteboom@uvt.nl
URL: <http://www.bartnooteboom.nl>

the focus will be on productive efficiency.¹ This paper attempts to give a systematic survey of different effects of scale in different service industries. We know a variety of effects of scale, mostly derived from manufacturing. The question is which of these apply to what kinds of services, and whether there are other effects of scale that apply to services rather than to manufacturing. The analysis is mostly limited to services provided by business firms, i.e. it does not go into public services, but medical services are included. It may be useful and feasible to extend the analysis further to public services, but that goes beyond the scope of the present paper.

For an understanding of productive efficiency in service industries, a first requirement is insight into the processes by which value is added. Porter's (1985) notion of the value chain has been useful for analyzing manufacturing processes. The question here is whether a similar construct, perhaps as a generalization of Porter's, may facilitate the analysis and categorization of the wide variety of service industries that we observe. Such a construct is expected to be useful for the analysis of sources of (in)efficiency, e.g. effects of scale, scope and experience. It is also useful for industrial marketing, to analyze where exactly one's goods or service impinges on a customer's value chain, to analyze its added value and obstacles for its implementation.²

A well-known characteristic of many services, often used as a defining characteristic (Gronroos 2000), is that product and production process are integrated, coinciding with time and place, and indeed the process, or part of it (the 'front office' of a service process), *is* the product, and that customers contribute to production. Another defining characteristic of services is that often the 'product', being abstract rather than physical, cannot be stocked. These features apply, for example, to the treatment of a patient in a medical facility, personal care, transportation, and recreation. However, while there are these fundamental differences between manufacturing and services, there are also similarities. Much manufacturing includes service elements (e.g., in marketing, logistics, product maintenance and support, ordering and billing, etc.). Some services may be produced 'industrially', in ways that are similar to manufacturing. Therefore, this survey will include manufacturing, with its elements of service, as well as service industries.

In this paper, the construct of a value chain is generalized and differentiated for the analysis of service value chains. This helps to categorize services into different generic types, and to assess what effects of scale occur where.

The generalized notion of a value chain compels us to ask useful questions such as follows:

- what, if anything, is 'processed', i.e. what is operated upon (natural deposits, raw materials, components, finished goods, plants or animals, bodies or minds of people, data or knowledge).

¹ For an analysis of effects of scale in innovation, see Nooteboom (1994).

² It was used for that purpose, in an analysis of opportunities for the application of information technology, in a course 'Entrepreneurship with information technology', at the University of Groningen, the Netherlands, in the 1990s.

- what does processing entail: some kind of transformation (physical, mental, informational), some kind of shift (in time or place), some kind of configuration (assembly, assortment composition), or some kind of saving or reduction (of complexity, risk).
- what value is added where, and what utility is generated: of form, function, design, time, place, health, well-being, pleasure or entertainment, information or understanding, safety or assurance.

Here, the working hypothesis is that answers to these questions will generate a useful categorization of services and a basis for detecting sources of (in)efficiency, in particular (dis)economies of scale.

The paper inevitably includes much that is already known about the separate industries. The added value here lies in finding a coherent conceptual system by which differences between industries are categorized and compared, and can more easily be analyzed.

The paper proceeds as follows. First, it gives a survey of known economies of scale, scope and experience. Second, it summarizes Porter's value chain for manufacturing, which serves as a point of departure, and presents the value chain for building construction, which in important aspects is similar to manufacturing. Third, it develops value chains for two non-manufacturing and non-service industries: agriculture and extraction, and analyzes corresponding economies of scale. Fourth, and this is the main objective of this paper, it turns to service industries, to develop corresponding value chains, and uses them to assign effects of scale to different types of service and different stages in their production.

2 Effects of scale, scope and experience

There are effects of scale, scope and experience in production, distribution, organization and networks. Effect of scale is defined as the effect on efficiency, measured as average cost per unit of production, of *more of the same*, i.e. of the volume of production per unit of time. It is mostly measured as sales volume, corrected for price and quality differences, or as added value, per unit of input of labor, capital or other production factors. Here, I will not discuss the measurement problems involved. Effect of scope is defined as the effect on efficiency of the range of *different* activities. Effect of experience is defined as the effect on efficiency of *cumulative* production over time.

One source of economy of scale is division of labor: because people specialize in a specific part of production they can perform their work more efficiently. The classic, paradigm case is Adam Smith's example of the pin factory, where people specialize in making the wire from which a pin is made, cutting off and sharpening the pin, attaching its head, and polishing the pin. With obvious differences in component activities, division of labor applies widely across industries, including service industries, as we will see.

Other economies of scale and scope are due to the presence of fixed ‘threshold’ costs. Threshold cost entails that productive capacity is not feasible or viable below some minimum level (is indivisible), such as a person, a team, or a machine. They may also be due to minimum setup costs of a facility for production, stocking, research, or service.

Some threshold costs have increased dramatically. This is the case, for example, in research and development of advanced technology and, as a result, in the development and production of many products that employ such technology, such as semi-conductors, airplanes, cars, consumer electronics. Formerly, products in consumer electronics could be made with low fixed costs of machinery and cheap labor owing to the manual assembly of components in boxes. With the onset of miniaturization, this practice of manual assembly was no longer viable, and it had to be replaced with the use of robots for assembly, which entails considerable threshold cost. On the other hand, some threshold costs have declined dramatically. This is the case, in particular, for computing, with the development of microcomputers and user-friendly software.

In fact, many economies of specialization can be reduced to economies in the utilization of threshold costs. Specialization of labor is feasible only if the volume of production is large enough to recoup the costs of investment in specialization: it takes a minimum of time and cost to train someone in a certain specialization, and sales must be large enough to recoup that threshold cost. In marketing also, there are substantial ‘threshold costs’, e.g. in distribution channels and brand name.

In distribution there are threshold costs of laying a pipe-line, (railway) track, road, of building a warehouse, building a station or airport, having an attendant at a service point, sending a truck along a certain route. For the development of a brand name one must invest heavily in advertising and other forms of promotion. The fixed cost of an advertisement on national TV is more easily recouped for a product that is marketed nationwide than for a local product. The need to market a specialized product worldwide, in order to gain sufficient economy of scale to compete, creates a need for a worldwide brand name, with correspondingly higher setup costs.

Further effects of scale in markets lie in ‘network externalities’, ‘bandwagon effects’ and ‘contagion effects’. In network externality (also called the ‘telephone effect’), the utility of some product depends on the number of other users who choose a product with the same, or compatible, technical standard. This can yield a ‘snowballing effect’ of a new standard that gains a head start before competing standards. This is a case of ‘first mover advantage’. In the ‘bandwagon effect’, customers copy each other’s choices. In the ‘snob effect’ it is the other way around: the choice of others is avoided.

In the diffusion of a new product, early adopters exert a demonstration effect (‘contagion’) upon potential later adopters. This yields the familiar S-shaped (‘logistic’) curve of diffusion. Such contagion may occur at a distance or, mostly with stronger effect, in direct contact, by ‘word of mouth’. Here, it matters who the source is. One is more inclined to follow people in similar

positions or roles (role equivalence or equivalence in network position; cf. Burt 1987).

Another quite different form of scale economy is due to the mathematical fact that the content of a container (e.g., reactor in the chemical industry) in the form of a sphere is proportional to the cube of the radius, and its surface is proportional to its square, while content yields production capacity and surface is connected with costs. The latter include costs of material, weight and air resistance, and hence transportation costs, costs of cleaning the surface, and costs of heat loss by radiation. As a result, the revenue per unit of cost increases as a function of volume of production. For a sphere it is easily calculated that a doubling of size (production volume) yields a 20% reduction of per unit cost.³ The effect has been called an ‘engineering’ economy of scale. It typically applies to process manufacturing (oil, chemicals, some pharmaceuticals, some food industries), where it has been called the ‘pots and pans effect’. The 20% reduction of cost at a doubling of volume is a familiar rule in those industries. However, the principle also applies, with some approximation, to the transport industry, as we shall see.

Economy of scope follows from different activities using the same resources, in different ways, or at different times, and thereby improving their utilization, or from different activities complementing each other in the utilization of resources.

The classic example of scope by joint utilization of resources is an orchard with sheep. For the sake of light and air, the trees need to be spaced out, and the resulting space between the trees is utilized for grazing sheep. Another example is the combination of different products in a distribution channel, to better utilize its capacity.

There are also several scale effects in transaction costs, i.e. costs of contact, contract and control (Nooteboom 1993). One is due to threshold costs of contact (search costs) and contract (evaluation and setting up a contract or other agreement), which weigh more heavily at small firm sizes. Furthermore, to the extent that knowledge is tacit there is less documentation as a basis for control, i.e. assessment and evaluation of competencies, needs, reliability, etc. Tacit knowledge also reduces absorptive capacity, and knowledge is especially tacit in small firms.

When knowledge is tacit, it tends to be self-evident, and not subject to rational criticism, until it has been sufficiently externalized in explicit knowledge. In other words, there is limited absorptive capacity until tacit knowledge is made explicit, and this is not always feasible. Note that thus tacit knowledge gives a double jeopardy: less information for evaluation and control, and a lesser basis for critical reflection and debate. The analysis explains the familiar phenomenon, identified in the small business literature, of how

³ If revenue for a spherical unit of production is proportional to volume v , which is proportional to the cube of radius r (r^3 , so that r is proportional to volume to the power $1/3$), and total cost is proportional to the square of the radius, then unit cost is proportional to $1/r$. The ratio of unit cost between two installations, with r_1 and r_2 , then is $r_2/r_1 = (v_2/v_1)^{1/3}$. Thus, If $v_2/v_1 = 2$, the unit cost ratio is 0.8.

difficult it is to obtain an audience for the adoption of innovations among small firms, and their inclination to reject it as ‘impractical’ (Nooteboom 1993).

There are also *diseconomies* of scale, particularly in organizations and networks. With n nodes of activity, the number of possible bilateral links between them is $n(n - 1)/2$, so that the number of possible linkages increases with the square of n . This can yield excessive costs of redundant linkages (Burt 1992), lack of flexibility and paralysis by excessive interaction, and confusion and deterioration of information by excess communication.

The classical solution is to reduce linkages by layers of hierarchy. A problem of hierarchy is that it may yield a loss of oversight and motivation. An alternative is a ‘hub and spoke’ structure with relations running via the central hub. A problem of a hub and spoke structure is that it may lead to informational overload in the hub.

3 Manufacturing and construction

As a point of departure and a benchmark, this paragraph summarizes the value chain and corresponding effects of scale for manufacturing industries.

3.1 Manufacturing

Porter’s (1985) scheme of the value chain, with a primary process of production and supporting staff services, has its paradigmatic application in manufacturing. There, the primary process is typically the transformation of physical inputs into a physical output, yielding a utility of form or function. In this paper I leave out the supporting services (personnel, legal, administrative, managerial) and focus on the primary process of production. Also, in addition to the flow of goods in the primary process I add a reverse flow of scrap for recycling and a reverse flow of information from the market, to feed new product development. The result is illustrated in Fig. 1. The depiction of the primary process is straightforward, with incoming logistics, including possible formation of stocks, the transformation and assembly process, outgoing

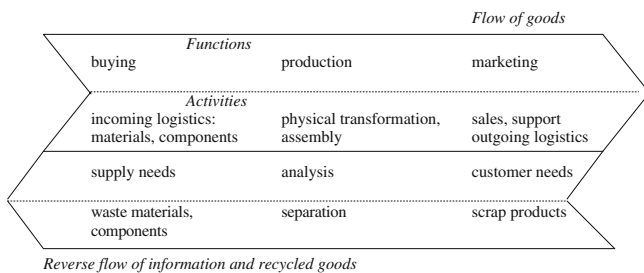


Fig. 1 Manufacturing value chain

logistics, and marketing, including sales and support in services such as installation, training and maintenance.

If marketing services are seen as an integrated part of the primary process of production, as was the case in Porter's original scheme, then if purchasing is seen as 'reverse marketing' (Leenders and Blenkhorn 1988), one might also include purchasing as part of the primary process, in view of close interaction with suppliers, in design, development, delivery, installation, support and maintenance of inputs. This is particularly the case under conditions of 'just in time' delivery, and 'Kanban' systems where supply of a single unit is 'drawn in' at the moment that it is demanded in the production process. It is also the case, in particular, under close cooperation with suppliers in 'co-makership' and 'co-design' (Lamming 1993). This move of including purchasing in the primary process is in contrast with Porter's original scheme, where it was part of support services.

Manufacturing processes have been categorized into single unit production (e.g., machinery), batch production (e.g., household goods), and flow or continuous production (e.g., in chemicals). While these differ considerably, in important ways, they do conform to the overall logic of Fig. 1 and will not be elaborated here.

Economies of scale arise, most of all, in the form of engineering economies in installations, particularly in the process industry, in economies of specialization, and in threshold effects in installations, machinery, instruments and supervision. Economies of scope arise in the leveraging of specialization, research, a variety of facilities, distribution channels, brand name, and a spreading of risk, with a portfolio of products that jointly optimize the utilization of resources or are complementary in other ways.

3.2 Construction

In construction, of houses, offices, hospitals, bridges etc., the primary process entails the collection, at a specific location, of workers, equipment, materials and components, for some transformation and mostly assembly into a structure, yielding a utility of form, function and place (location). The units processed are materials and components. The process is largely location-bound: most activities take place on site. Unlike most manufacturing, construction is project based: management, workers, installations and materials are configured for a specific project at a specific place. The stages of the process are collecting workers, equipment, materials and components, then transformation and assembly, and finally return of equipment and workers. This is illustrated in Fig. 2.

There are economies of scale in threshold costs of equipment (e.g., cranes), transportation of the many components of the process (economies in transport will be analyzed later), and specialization of skills. The construction process is highly systemic, with a great number of elements that are densely and tightly connected in specification and timing. This complicates the achievement of efficiency. There are large experience effects in coordination.

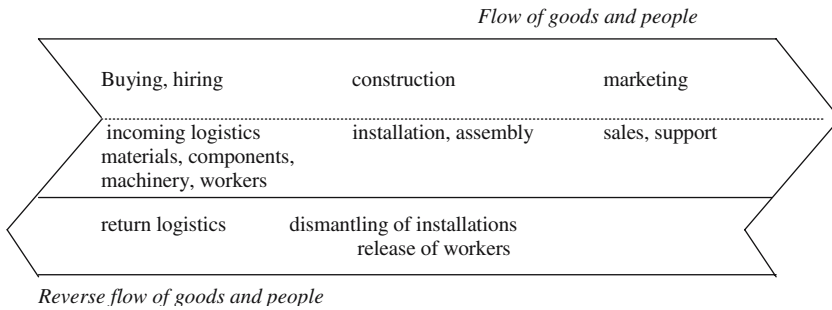


Fig. 2 Construction

Systemic production is not unique for construction and also obtains, for example, in the assembly of cars. There also, many tightly connected elements are provided by a great diversity of suppliers. An important difference is that in car assembly there is continuity in an ongoing flow of production, coordinated by a single and well-identified end producer, while in construction production is on a mostly ad hoc project basis, with the lack of a clear coordinating actor. As a result there is great inefficiency through a mismatch in forms (measures, shapes, technical standards, materials, colors), with elements not matching the structures in which they need to fit, and in timing, with a plumber arriving after the walls have been plastered, destroying the plastering for cutting ducts for piping. There are estimates that up to 20% of total costs arise from such mismatches. Thus, there is a pressing need to coordinate activities. This is complicated by differences in perspectives and mind frames, and in interests, yielding lack of understanding and rivalry between architects, developers, surveyors, builders, subcontractors, housing agencies, and users. As a result, there is pressure towards integration, with one of the many actors serving as an integrator (Cacciatori and Jacobides 2005).

Efficiency is also hindered by an orientation, in contracting, towards the lowest price. This hinders open book contracting that would be needed for proper coordination: parties need to keep their costs and processes secret to maintain bargaining position, which hinders mutual adjustment for quality and efficiency, while the low price is not in fact achieved because profit is recouped in redressing unforeseen, uncontracted problems and mismatches. In the car industry this problem is prevented by price minus costing that guarantees a profit for suppliers, which makes them accept open book contracting. In the building industry this is prevented by a lack of continuity of buyer–supplier relationships, and an ingrained mentality that is oriented towards price rather than quality (Nooteboom 1998).

4 Natural resources

This paragraph analyzes industries that fall neither under manufacturing nor under services, and operate on natural resources: agriculture and extraction.

4.1 Agriculture

Agriculture resembles manufacturing in that the process is one of physical transformation, yielding a utility of function and/or form. However, it is deeply different in several respects. The units processed are forms of life, the process of transformation is one of natural growth rather than artifice, and production entails the facilitation of that process rather than its execution. The process is bound to a location (soil), with corresponding vicissitudes of climate. The process is characterized by what Aristotle called ‘entelechy’: potential form is not designed by man but predetermined by nature and locked up in a genotype. Agriculture will become more like manufacturing under genetic engineering. The stages of an agricultural value chain are seeding, feeding, harvesting and culling. This is illustrated in Fig. 3. There are, of course, important differences between growing plants and raising livestock, but both satisfy the basic logic of Fig. 3.

There are economies of scale in threshold and setup costs of a farm and in equipment (tractors, harvesters, milking installations, etc.), in logistics (utilization of transport for moving materials and products), and in specialized knowledge and skill needed for different growth processes. There is also an effect of scale in guiding a herd of livestock: the minimum possible capacity is one person, regardless of the size of the herd. There are important economies of scope in combining different forms of life, to spread risks of weather, disease, and price fluctuations in markets, to rotate crops and utilize different seasons for different crops, and in combining crops and livestock (for grazing and the provision of manure). Small farms are faced with the dilemma of on the one hand having to limit the scope of activities in order to achieve a minimum efficient scale per form of life, and on the other hand wanting to spread risks across a wider scope of activities. Thus it is particularly important for smaller farmers to reduce effects of scale by collaboration, often in the form of cooperatives, to provide support in specialist knowledge and in renting or lending machinery.

4.2 Extraction

Extraction industries resemble agriculture in that they operate on natural resources. They delve raw materials from organic or mineral deposits, such as iron ore, bauxite (for aluminum), coal, oil, etc. They entail two very

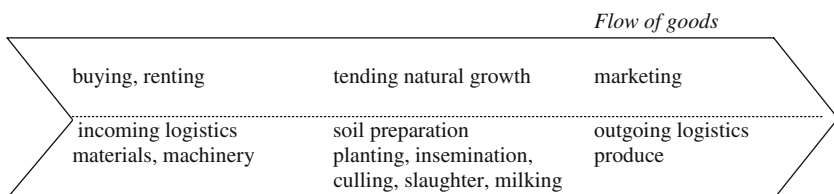


Fig. 3 Agriculture

different kinds of activities: exploring the location of fields and the actual extraction. Here I focus on the latter. The process entails extraction from soil or water, and separation, cleaning, and transportation of raw materials. Like manufacturing, and more particularly like construction, it has incoming and outgoing logistics. The stages of the process are collecting workers and equipment, next extraction, separation and cleaning, and finally transportation of materials, and return of equipment and people. This is illustrated in Fig. 4.

There are large economies of scale in the setup and utilization of large and often expensive equipment (e.g., oil rigs), safety and security (guards), housing and health of personnel at often remote locations, and in transportation to and from a site. There may be engineering economies of scale in equipment. There are economies of scope in the multiple usage of equipment, know-how, and supporting services, for different types and grades of raw materials and different locations. Owing to the inaccurate estimates of the size and accessibility of deposits individual projects may be risky, and there are economies of scope in spreading the risk across multiple projects at different locations. As for agricultural products, raw material markets are subject to fluctuations of price, and there are economies of scope in spreading risks across different kinds of deposits.

5 Distributive services

This paragraph analyzes service industries that yield supply and distribution, mostly of goods, yielding utilities of time, place and assortment.

5.1 Utilities

Utilities entail the supply of gas, water, electricity, telephone service, internet access and cable TV, with a utility of time, place and reliability of supply, repair, and billing. The primary process is divided into two quite different

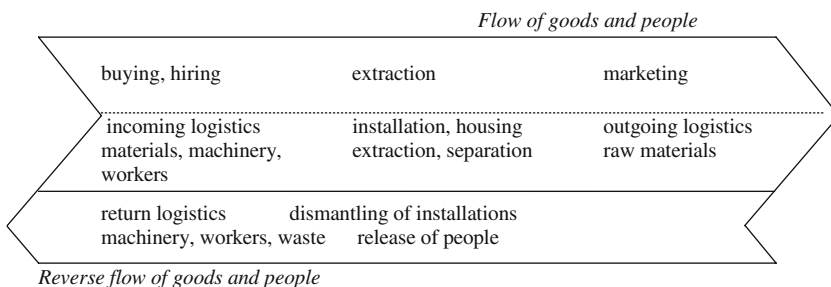


Fig. 4 Extraction

processes of production and delivery. Production is a part of manufacturing, resembling other process industries (such as chemicals), analyzed earlier, and will not be further considered here. Delivery entails a flow of materials (gas, water, electricity) or information through a network infrastructure of pipes, cables or telecommunication.

In delivery there are large economies of scale in equipment (ducts, propulsion, switching, compression, amplification) and in the utilization of network capacity. There are network externalities in technical standards. There are economies of scope in 'peak shaving', back-up for failures, in the connections between elements of the network and in different uses of a network (e.g., the joint use of cables for telephone, internet services and TV). Costs of the network infrastructure are highly 'sunk', i.e. the network often is highly dedicated (with the exception of cables for telecommunication), with limited alternative uses. This yields an incentive to leave network structure and its utilization under public control. The provision of services on the network, on the other hand, is increasingly subject to privatization. As a result there is pressure for unbundling infrastructure and delivery. However, that still requires public regulation and control of the conditions for access to the network for competing suppliers.

Privatization of supply yields incentives for suppliers to utilize economies of scope in providing additional services that utilize know-how, competencies (reliable supply, billing), customer base, and brand name. For example, energy suppliers are stimulated to include the supply of other utilities (water, telephone, TV), and to diversify into activities such as energy management, including the supply and maintenance of equipment (for cooking, heating, cooling), home security systems (around the theme of reliability), 'intelligent homes' (with IT-enhanced equipment communicating among each other), monitoring and emergency systems for the elderly remaining at home, and even home health care. Such diversification of competing suppliers into different portfolios of activities is likely to yield increased dependence of consumers on integrated suppliers, reduced price competition between them, and increased concentration. This raises doubts whether in the end allocative efficiency by competition is indeed enhanced.

5.2 Transportation

In transportation the basic process is that objects or people undergo a shift of place, at a certain speed, yielding a utility of place and time. Further value is added in safety of travel, range and reliability of travel schedules, and connections with ongoing travel. For personal transport there is also comfort of travel (space, seats, beds), catering, personal attention, pre-departure and post-arrival services (reception, guidance, parking). People and objects may be moved in parallel, as when luggage and catering are processed in airline traffic. This value chain is illustrated in Fig. 5.

Personal transport is one of the services that most satisfies the characteristic that the product is the process, and production and consumption coincide with

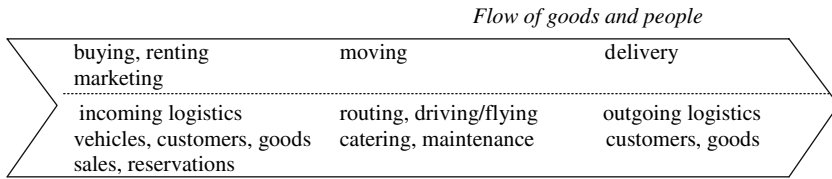


Fig. 5 Transport

time and place. To a large extent, transportation is an ‘experience product’, whose quality is mostly judged during consumption. Value is added in different stages of the process:

- Pre-departure: parking, reception (structurally comparable to incoming logistics in manufacturing), check-in, luggage, security checks, waiting for departure (comparable to stocking in manufacturing), with corresponding facilities, catering and perhaps shopping, and perhaps entertainment (such as casinos in airports, and, since recently, an exhibition of Rembrandt paintings at Schiphol airport, and, as recently announced, a small golf course).
- Transport: seating, catering, entertainment (music, films), safety procedures.
- Post-arrival: docking, check-out, luggage retrieval, connecting transportation, accommodation, lost luggage, customs, complaint procedures.
- Particularly in goods transport, there is added value in the informational ‘tracking and tracing’ of goods in transport.

Operating an airplane, bus or truck combines two effects of scale. First there is the engineering effect of scale due to the mathematical ratio of surface to volume. A large plane, for example, as bulbous as it can get, subject to constraints of properties for flying, as in a jumbo jet, offers more passenger space (content), relative to costs of materials and costs of movement (surface). Second, there is a threshold effect due to the need for a minimal crew to fly the plane, regardless of plane size. Another effect of scale is associated with the fact that with more loads the utilization (payload) of space increases. Yet another lies in the servicing of vehicles at different locations: one needs a large volume to efficiently utilize service capacity. Yet another lies in booking systems, via agencies or websites.

In the airline industry, an important effect of scope arises in connecting flights for the convenience of travelers, in ‘code sharing’. Having a wide portfolio of flights, or an alliance, allows one to offer better connections. Another economy of scope lies in combining flights between different destinations to reduce the number of flights, in ‘hub and spoke’ configurations. Between n different destinations there are $n(n - 1)/2$ possible flights, while connections directed through a central hub yield only $n + 1$ flights. Also, while in that system the average distance of flights increases, the number of flights is lower, so that the number of passengers per flight increases, allowing for economies of both scale of plane size and payload.

All these effects of scale and scope converge in pressures on airlines (and trucking companies), to engage in mergers and acquisitions or close alliances.

5.3 Delivery points

Delivery points are a generalization of retailing. They allow people to deposit, collect or consume goods or information at a certain location (a geographic location or a location on the web). This might be interpreted so broadly as to include retailing, also retail banking, cafes/restaurants/hotels, travel bureaus, information or support services (weather forecasts, lost and found, computer service desks, etc.), municipal and administrative services (e.g., information on zoning, social support, social security, passports), emergency calls (police, fire brigade, ambulance, etc.). It might even be seen to include recreational facilities (theatres, cinemas, sports, amusement parks, etc.), services of personal care (e.g., hairdressers), health (doctors, hospitals), and education. However, I draw the line here between more narrowly distributional services, with an emphasis on utility of time, place and assortment, and services that affect the physical or mental state or well-being of people (recreation, health, education), which will be analyzed later. Nevertheless, the boundary is fuzzy. In particular, cafes/restaurants/hotels border on recreational services, and might indeed be combined with them. Retailing also may be seen as increasingly developing into a form of recreation. Conversely, services that affect physical or mental well-being or capacities of people share important features with distributional delivery points, as will be discussed later.

For delivery points, a wide assortment of goods and/or services allows for ‘one-stop-shopping’. A crucial feature of delivery points is that efficiency depends greatly on the utilization of service capacity while this is primarily up to customers, not producers. The value chain is illustrated in Fig. 6.

Thresholds arise in the form of the presence of the minimum of one attendant during opening time (unless the service is fully automated). This also applies to a teacher before a class, or a nurse on duty in a hospital ward. Threshold staff or facilities are under-utilized during low levels of demand

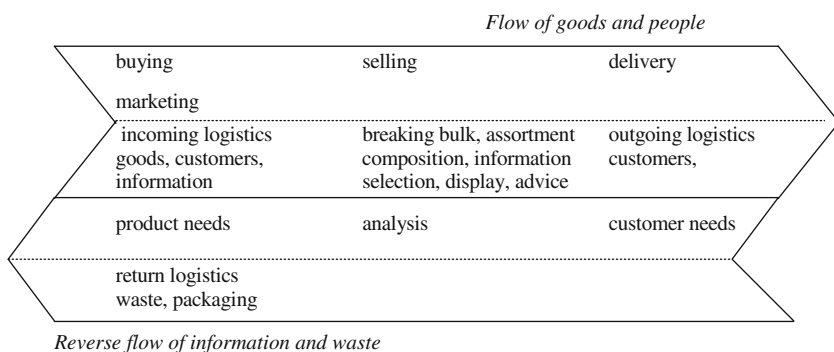


Fig. 6 Delivery points

(low customer visits, small classes, few patients), and this yields a scale effect. At higher levels of demand, when threshold labor is not enough to support an adequate service level, additional labor may be added part time, only during peak hours.

Nootboom (1982) argued that for this type of production, which unlike manufacturing does not entail any transformation of form or function, an appropriate tool for the analysis of efficiency, in the utilization of service capacity, is queuing theory, to deduce capacity utilization from arrival patterns of customers. The conclusion was that a linear curve of labor costs applies, with its intercept equal to the sum of opening times of independently staffed service points (threshold costs). The straight line with that intercept approximates the cost line when average queuing time of customers is about equal to average service time per customer. This entails an effect of scale in the 'threshold' capacity becoming better utilized as the number of customer visits increases.

An important implication is that an increase of opening times of shops, as a result of the liberalization of shopping hours, in a sequence of European countries, increased the intercept of the line (in an upward shift), and thereby intensified the effect of scale, leading to an accelerated decline of small relative to large shops. As a result, an increase of the utility of (opening) time led to a decrease of the utility of place, due to increase of scale and hence an increase of average distance from consumers to shops. This effect was selective: it applied less to the more prosperous and mobile consumers, who also benefited most from an extension of shopping hours, and more to less prosperous and older, less mobile consumers, especially in older city quarters and small villages, with important secondary effects. Small villages were emptied more rapidly of young families, which caused schools to leave villages, which contributed to an accelerated decline of public transport. The larger distance to shops and the decline of public transport contributed to the poorer elderly moving to pensioners' homes earlier.

The increase of scale of shops was also considerably enhanced by the invention of self-service. The introduction of the principle of self-service entailed a reconfiguration of roles in the shopping process, particularly some reversal of roles between retailer and customer.

In a service shop the customer is stationary, waiting at a service counter, while the shop attendant is mobile, moving up and down the shop to collect goods. In self-service retailing, this is reversed. The cashier is stationary at the check-out, and the customer moves up and down aisles in the shop. The switch entails a substitution of customer labor and shop space for supplier labor: multiple consumers moving about the shop requires more space. The shift to self-service constituted a major organizational innovation. Above all, it eliminated a fundamental obstacle to a larger range of goods that were obtained in a service shop. There, for an extended range of goods, shop attendants would have to walk greater distances to collect goods, yielding inordinate waiting times of customers at the service counter.

In other words, the newly emerging utility of a large assortment was enabled by the invention of self-service. On the demand side there was not only an increased demand for bulk shopping, but also a decreased need for product advice from the attendant, because due to increased prosperity and widespread mass advertising, consumers became more knowledgeable and confident in selecting a wide range of products. This was an enabling factor for self-service. The effect of self-service is reflected in the coincidence of the increase of store size with the advent of self-service (Henksmeier 1960; Dawson 2000).

There are clear economies of scope in adding services that improve the utilization of service capacity, particularly when additional activities can take place at times when few customers are present. In retailing, facilities are better utilized with the addition of products. In travel agencies the sale of insurance may be added to the booking of travel and accommodation. In municipal services it pays to combine e.g. passports, drivers' and other licenses, parking fees, and assorted information services.

In retailing, there are also a variety of scale effects beyond shop size, on the level of the retail enterprise, in the form of integrated chain stores, and cooperatives and voluntary chain stores (in fact a form of franchising) of independent retailers.

One, arguably the most important, effect is that of discounts for purchasing in bulk from manufacturers, based on economies of scale in large scale manufacturing, bulk ordering (more goods per order), transport (better utilization of transport capacity), and stocking. Beyond mere efficiencies, however, there are also issues of negotiating power. Among manufacturers there is competition for shelf space in retailing, and it is the retailers that control that space. Especially for new product introductions, nationwide retail chains can negotiate hefty price reductions for introducing the product to the market. Furthermore, large chains have the volume to integrate backwards into manufacturing, with 'own labels'. The threat of that can contribute to further price rebates from established manufacturers. Some manufacturers are tempted to make their production capacity available to produce such retail labels, sometimes next to their own labels, to utilize excess capacity.

Other effects of scale and scope lie in obtaining finance, spreading risks among different locations and types of shop, ability to initiate and sit out price wars, negotiations for attractive shopping locations and conditions with municipalities and shopping mall developers and builders, nationwide advertising to build up and maintain a retail brand name, and in human resources. In the latter, there are economies in training, hiring and selection, and in the offer of longer career paths upwards in the hierarchy of jobs in shops, regional management, staff jobs, and central management. The increase of market share of the ten largest firms from just under 28% in 1992 to just over 36% in 1999, in Europe (Dawson 2000), appears to confirm enterprise economies of scale. Kuwahara (1997) showed evidence of increasing concentration in Japan.

6 Personal services

This paragraph analyzes personal services, in health and other care (hair-dressing, cosmetics, etc.), recreation, professional knowledge services (e.g., consultants, accountants, etc.) and education. These all affect either the body or the mind of people, or both, and yield utilities of physical or mental well-being or capacity.

6.1 Personal care

First, I group together health care, other personal care, and recreation, in spite of the many differences between them, because in basic value chain logic they appear to be similar. In all these cases, primarily people (patients, customers) flow through the system for some kind of treatment that affect body, mind or both. There is incoming logistics of customers, with a need for easy access and parking, and of materials, instruments or installations, and workers. There often are ancillary services of catering and accommodation. There is outgoing logistics of customers after they are released from treatment. In addition to primary utilities of physical and/or mental well-being, there is often an important utility of safety and reliability. The value chain is illustrated in Fig. 7.

Of all services, together with transportation and education these services of personal care most have the property that the product is the process, customers contribute to production, and production and consumption coincide in time and place. However, there is a sizeable ‘back-office’ of activities ‘behind the stage’, in personnel training, facility management, and administration.

For customers the product is always an ‘experience product’, whose quality is judged during consumption. To a large extent, in health care, and education, to be analyzed later, the product is also a ‘credence product’, whose quality is difficult for the patient to judge even after treatment. This makes patients highly vulnerable and dependent on reputation, brand, and supervision, by public or professional bodies, of professional conduct. Reputation entails economies of time and experience, and there are economies of scale in building a brand name by advertisement.

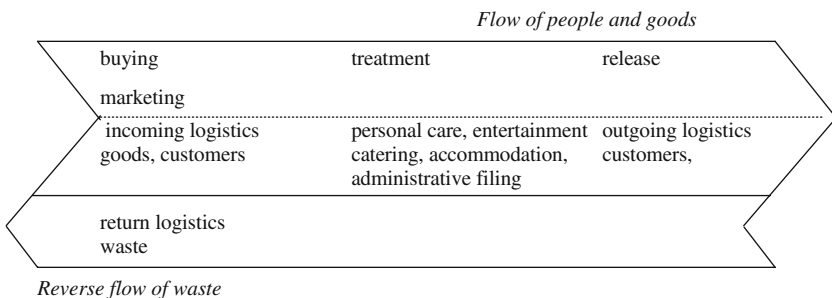


Fig. 7 Personal care/recreation

Both engineering economies and threshold economies of scale may apply to installations, machinery and instruments (e.g., high-cost medical apparatus such as a brain scanner). There are clear economies of specialization and division of labor. However, in hospitals there are also serious diseconomies in the compartmentalization between stages of treatment, between specialists and departments, which yield delays, duplication of treatment and communication, and mismatches and miscommunication that can have serious consequences.

There are considerable economies of scope in leveraging facilities by the addition of ancillary services, such as a wider range of medical services in a hospital, and addition of catering and accommodation in recreation.

These services have an important feature in common with the delivery points discussed earlier, in the (dis)economies due to threshold costs of personnel that must be available during opening times of a service, and whose utilization depends on stochastic arrival times of customers. However, in contrast with e.g. retailing here demand can more easily be spread across time to match service capacity, by means of systems of appointments (in care) or reservations (in entertainment). Nevertheless, there are peaks and troughs in demand, and the use of part-time labor is important to 'shave peaks', to the extent that the flow of customers cannot be strictly controlled (e.g., in recreation). Earlier I noted how close some delivery points, such as cafes/restaurants/hotels, but also increasingly retailing, come to recreation and entertainment.

6.2 Knowledge

Knowledge services include education, training, and the provision of information or advice (consultants, accountants, notaries, etc.). They have in common that they yield utility in enhancing the knowledge, insight, mental capacity or skill of customers. The value chain is illustrated in Fig. 8.

There is a sizeable 'back-office' activity in data processing, administration, training of personnel, and research that feeds the knowledge service. There are economies of scope in the extension of services to better utilize available facilities and personnel in times of slack.

While knowledge services are to some extent experience products, whose quality is determined during consumption, they are largely credence products,

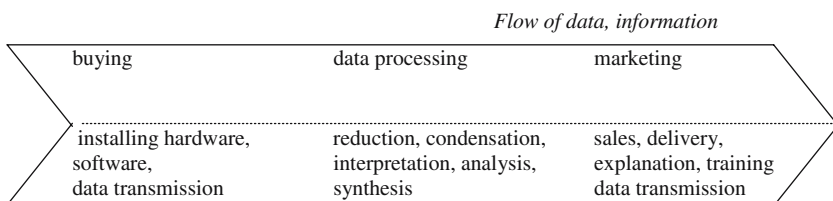


Fig. 8 Knowledge

whose quality can only be judged much later, when applying knowledge in practice, or not at all. Thus, as in the case of the personal services discussed earlier, reputation, brand name and institutionalized control of professional performance are of great importance. There are economies of time and experience in building reputation, and economies of scale in building a brand name by advertisement.

There are clear economies of specialization, which, however, may also degenerate into fragmentation and lack of coherence. There is a large threshold effect in education, of a teacher serving a class of students, whether that is large or small. Hence there is a trade-off between specialization and class size, and a pressure towards large institutions to allow for a combination of large class size and specialization. However, here one soon runs into diseconomies of scale involved in deliberation needed for mutual adjustment, to achieve some coherence, given the fact that with n teachers there are $n(n-1)/2$ possible bilateral contacts for deliberation. To avoid explosion of communication, some hierarchy of levels and departments may be needed, but this may reduce the motivation, initiative, and the autonomy and room for experimentation and variation needed for inventive intellectual activity. Thus, there is a trade-off between size, to allow for specialization and large class size, and an overload of communication and demotivation.

7 Finance and insurance

Finally, there are services in finance and insurance. The value chains of these activities are sufficiently similar to be put together in a single scheme, as illustrated in Fig. 9. Finance and insurance are also connected in that proceeds from insurance premiums are invested for profits, and financial risks may in turn be insured. This similarity and connection explains why finance and insurance have increasingly been integrated in financial corporations. Of course, banks have retail outlets, which would fall under the category of

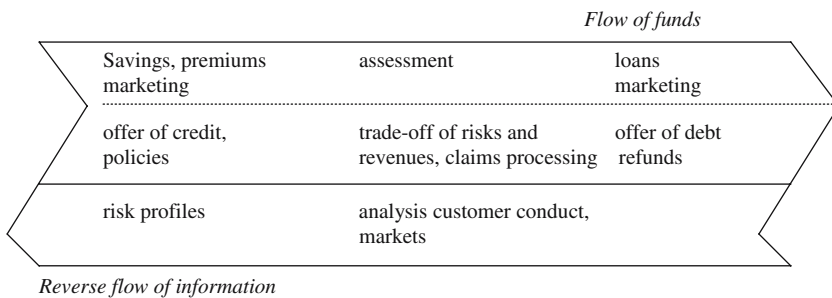


Fig. 9 Finance, insurance

delivery points, as discussed earlier. Here the analysis turns to the underlying activities, in the 'back office'.

This industry is similar to knowledge industries, in that the analysis of data, here about financial markets, revenues and risks, is a core activity. In finance a science has developed concerning asset pricing and the valuation of options. In insurance there is a science concerning 'adverse selection' (how people with the worst risks have the greatest interest in insurance) and 'moral hazard' (how once insured people become careless). These analyses share an underlying quantitative science of probabilities and statistics.

In the value chain, there is an incoming flow of information, to compute risks and revenues, and of funds from saving accounts and insurance premiums. This entails marketing, at the incoming side of the chain, in the search of customers, and in the development and sale of a portfolio of products for saving and insurance. Incoming funds from savings and insurance premiums are invested in an outgoing stream of loans, and insurance claims processing yields an outgoing stream of refunds. Here there is marketing, at the outgoing side of the chain, in the development and sale of loans, mortgages etc. There is an important reverse flow of information from markets, concerning revenues and risks and related customer conduct, to feed the input of data for improved assessment of revenues and risks.

Finance and insurance both yield utility of funds, diversification of risk, and assurance. To some extent these services have characteristics of an experience good, in the helpfulness and the speed of service experienced. Consider, for example, the speed of claims processing in insurance. However, they are also to a large extent credence products. It is difficult for most customers to assess whether correct advice and implementation is given. A large part of utility is security against financial default. This is so large that a central bank is required to monitor bank practices. For remaining uncertainties, reputation and brand name are of great importance.

There is economy of scale due to threshold costs in offices and in technical facilities, such as computer networks. There are economies of scale and experience in the building and maintenance of reputation and brand name. There are economies of scope in risk diversification across a portfolio of investments and insurance risks. There is economy of scope also in the utilization of facilities and brand name with a range of activities and products. There are economies of specialization, e.g. in knowledge concerning different markets, industries, and kinds of insurance risks. One must be large to achieve a wide portfolio of specialties, in order to spread risks and achieve economies of scope, as well as a sufficient volume in each specialty to achieve minimum efficient scale. Under globalization, there is pressure, particularly in business markets, to offer worldwide services, either by expansion or by alliances. An alternative is to focus only on a specialized area, in a niche market, with more dedicated service, yielding a lesser degree of risk diversification but a higher margin.

8 Conclusion

A generalization of Porter's notion of a value chain is useful for the analysis of service industries. It helps to detect sources of (in)efficiency, e.g. due to (dis)economies of scale, scope and experience. It can also be useful for suppliers to detect opportunities and obstacles for marketing their products to these industries. The generalization entails that the flows and the physical transformation of goods that are characteristic of manufacturing are generalized into flows and transformation of data and flows and transformation of the physical and mental condition of people that are characteristic of many service industries. Utility is generalized from utilities of form and function of goods, characteristic of manufacturing, to utilities of time, place, convenience, speed, entertainment, physical and mental well-being, knowledge and mental capacity, and management of funds and risks. A survey is given in Table 1.

Table 1 summarizes key differences between industries, concerning what is processed, characteristics of the process, and types of utility generated. The characterization of the process also includes the time aspect, i.e. whether the process is ongoing (as in manufacturing), intermittent in the sense of being project based (as in construction and extraction industries), or 'episodic'. By 'episodic' I mean that production entails spurts of service, using a generic service capacity in a dedicated fashion for individual customers (as in transport, entertainment, education, and, perhaps increasingly, in finance and insurance).

Table 1 Survey of industry characteristics

Industry	What is processed	Type of process	Utility
Manufacturing	Materials, components	Physical transformation assembly; ongoing	Form, function
Construction	Materials, components	Assembly, project based	Form, function, place
Agriculture	Forms of life	Natural growth; seasonal	Mostly nutrient
Extraction	Natural resources	Extraction; project based	Function
Utilities	Natural resources, data	Transmission; ongoing	Place, time, reliability
Transportation	Goods, people	Moving; episodic	Place, time, safety, comfort
Delivery points	Goods, information	Assortment composition, place, display, ongoing	Time, assortment, atmosphere
Personal care	People	Transformation of body and or mind; episodic	Physical and mental well-being
Knowledge	Information, people	Addition to or transformation of mind	Knowledge, mental capacity
Finance	Information, funds	Assigning debt and credit; ongoing/episodic	Capital, interest
Insurance	Information, funds	Assessing risks and claims; ongoing/episodic	Security

Of course, for this wide survey across industries many details, and differences across conditions, e.g. in different countries, had to be left out. For practical purposes many of such details and specificities would have to be added. Comparison across industries shows up both differences and similarities. The similarities between different industries yield boundary problems, of whether a given case belongs in the one or the other category. For example, the cafe/restaurant/hotel industry has features in common with both retailing (as delivery points of goods) and entertainment. Innovations can consist of novel combinations of elements from different categories. For example, some innovation in retailing consists of bringing in features of entertainment. Entertainment is also increasingly brought into transport, with films in planes and buses, and casinos and even golf facilities at airports. Even so, the scheme offered here may be useful precisely to explore opportunities for such innovative boundary crossing, and for assessing implications for (in)efficiencies of scale, scope and experience in such novel combinations.

References

- Burt R (1987) Social contagion and innovation: cohesion versus structural equivalence. *Am J Sociol* 92:1287–1335
- Burt R (1992). *Structural holes: the social structure of competition*. Harvard University Press, Cambridge
- Cacciatori E, Jacobides MG (2005) The dynamic limits of specialization: Vertical integration reconsidered. *Organ Stud* 26:1851–1884
- Dawson JA (2000) Retailing at century end: some challenges for management and Research. *Int Rev Retail Distrib Consum Res* 10(2):119–148
- Gronroos C (2000) *Service management and marketing*. Wiley, New York
- Henksmeier KH (1960) The economic performance of self service in Europe. OECD, Paris
- Kuwahara H (1997) Concentration and productivity in the retail trade in Japan. *Int Rev Retail Distrib Consum Res* 7(2):109–124
- Lamming R (1993) *Beyond Partnership*. Prentice Hall, New York
- Leenders MR, Blenkhorn DL (1988) *Reverse marketing; The new buyer–supplier relationship*. The Free Press, New York
- Nooteboom B (1982) A new theory of retailing costs. *Eur Econ Rev* 17:163–186
- Nooteboom B (1993) Firm size effects on transaction costs. *Small Bus Econ* 5:283–295
- Nooteboom B (1994) Innovation and diffusion in small business: theory and empirical evidence. *Small Bus Econ* 6:327–347
- Nooteboom B (1998) Cost, quality and learning based governance of transactions. In: Colombo M (ed) *The changing boundaries of the firm*. Routledge, London, pp. 187–208
- Porter ME (1985) *Competitive advantage; creating and sustaining superior performance*. The Free Press, New York